BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e37424
doi: 10.3897/biss.3.37424

OPEN ACCESS

Conference Abstract

SEINet: A Centralized Specimen Resource Managed by a Distributed Network of Researchers

Edward Gilbert‡, Corinna Gries§, Nico Franz‡, Landrum Leslie R.‡, Thomas H. Nash III§

‡ Arizona State University, Tempe, United States of America
§ University of Wisconsin, Madison, Madison, United States of America

Corresponding author: Edward Gilbert (egbot@asu.edu)

Received: 17 Jun 2019 | Published: 26 Jun 2019

Citation: Gilbert E, Gries C, Franz N, Leslie R. L, Nash III T (2019) SEINet: A Centralized Specimen Resource Managed by a Distributed Network of Researchers. Biodiversity Information Science and Standards 3: e37424. https://doi.org/10.3897/biss.3.37424

Abstract

The SEINet Portal Network has a complex social and development history spanning nearly two decades. Initially established as a basic online search engine for a select handful of biological collections curated within the southwestern United States, SEINet has since matured into a biodiversity data network incorporating more than 330 institutions and 1,900 individual data contributors. Participating institutions manage and publish over 14 million specimen records, 215,000 observations, and 8 million images. Approximately 70% of the collections make use of the data portal as their primary "live" specimen management platform. The SEINet interface now supports 13 regional data portals distributed across the United States and northern Mexico (http://symbiota.org/docs/seinet/). Through many collaborative efforts, it has matured into a tool for biodiversity data exploration, which includes species inventories, interactive identification keys, specimen and field images, taxonomic information, species distribution maps, and taxonomic descriptions.

SEINet's initial developmental goals were to construct a read-only interface that integrated specimen records harvested from a handful of distributed natural history databases. Intermittent network conductivity and inconsistent data exchange protocols frequently restricted data persistence. National funding opportunities supported a complete redesign towards the development of a centralized data cache model with periodic "snapshot" updates from original data sources. A service-based management infrastructure was

© Gilbert E et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

integrated into the interface to mobilize small- to medium-sized collections (<1 million specimen records) that commonly lack consistent infrastructure and technical expertise to maintain a standard compliant specimen database. These developments were the precursors to the Symbiota software project (Gries et al. 2014). Through further development of Symbiota, SEINet transformed into a robust specimen management system specifically geared toward specimen digitization with features including data entry from label images, harvesting data from specimen duplicates, batch georeferencing, data validation and cleaning, generating progress reports, and additional tools to improve the efficiency of the digitization process. The central developmental paradigm focused on data mobilization through the production of:

1. a versatile import module capable of ingesting a diverse range of data structures,
2. a robust toolkit to assist in digitizing and managing specimen data and images, and
3. a Darwin Core Archive (DwC-A) compliant data publishing and export toolkit to facilitate data distribution to global aggregators such as Global Biodiversity Information Facility (GBIF) and iDigBio.

User interfaces consist of a decentralized network of regional data portals, all connecting to a centralized shared data source. Each of the 13 data portals are configured to present a regional perspective specifically tailored to represent the needs of the local research community. This infrastructure has supported the formation of regional consortia, who provide network support to aid local institutions in digitizing and publishing their collections within the network. The community-based infrastructure creates a sense of ownership – perhaps even good-natured competition – by the data providers and provides extra incentive to improve data quality and expand the network.

Certain areas of development remain challenging in spite of the project's overall success. For instance, data managers continuously struggle to maintain a current local taxonomic thesaurus used for name validation, data cleaning, and to resolve taxonomic discrepancies commonly encountered when integrating collection datasets. We will discuss the successes and challenges associated with the long-term sustainability model and explore potential future paths for SEINet that support the long-term goal of maintaining a data provider that is in full compliance with the FAIR use principles of making the datasets findable, accessible, interoperable, and reusable (Wilkinson et al. 2016).

Keywords

natural history collections, biodiversity data, specimen digitization, open access, SEINet, Symbiota

Presenting author

Edward E. Gilbert

Presented at

Biodiversity_Next 2019

References

- Gries C, Gilbert E, Franz N (2014) Symbiota – A virtual platform for creating voucher-based biodiversity information communities. Biodiversity Data Journal 2 https://doi.org/10.3897/bdj.2.e1114
- Wilkinson M, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J, da Silva Santos LB, Bourne P, Bouwman J, Brookes A, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo C, Finkers R, Gonzalez-Beltran A, Gray AG, Groth P, Goble C, Grethe J, Heringa J, 't Hoen PC, Hooft R, Kuhn T, Kok R, Kok J, Lusher S, Martone M, Mons A, Packer A, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone S, Schultes E, Sengstag T, Slater T, Strawn G, Swertz M, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3 https://doi.org/10.1038/sdata.2016.18